DOCUMENT RESUME

ED 055 740 RE 003 836

AUTHOR	Tuinman, J. Jaap; And Others
TITLE	Increases in Test Scores as a Function of Material Rewards.
INSTITUTION	Indiana Univ., Bloomington. Inst. for Child Study.
PUB DATE	71
NOTE	25p.
AVAILABLE FROM	Director, Institute for Child Study, Indiana University, Bloomington, Indiana 47401
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Educational Research; *Junior High School Students; Motivation; *Reading Comprehension; Reading Improvement; Reading Research; *Rewards; *Test Results

ABSTRACT

From the entire population (N=341) of grades 7 and 8 in a rural Indiana junior high school, 160 subjects were randomly selected and assigned to the experimental and the control groups. Form A of the Nelson Reading Test was administered twice with a 4-week interval. While the control group was told only that the post-test was given to measure how much they had learned since the last testing, the experimental group was told that those making any performance improvement would win material rewards. Those making the most increase received one of six transistor radios; others received a university sweatshirt or a 25-cent candy bar. Results were analyzed using analysis of covariance. Two major dependent variables were the total number correct (NC) and the total number of items attempted (NA). Post-test means were adjusted for pretest differences between the two groups. The effect of the reward treatment was found to be significant in all instances on the NC variable ($p<.01$) and on the NA variable ($p<.001$). Results for the vocabulary and comprehension subtest were also reported. It was indicated that the experimental group outperformed the control group by more than 3 months in total means. Tables and references are included. (AW)

Increases in Test Scores as a Function of Material Rewards

J. Jaap Tuinman

Roger Farr

B. Elgit Blanton

Institute For Child Study

School of Education Indiana University

Institute For Child Study

School of Education Indiana University

Increases in Test Scores as a Function of Material Rewards

J. Jaap Tuinman

Roger Farr

B. Elgit Blanton

The Institute for Child Study is an interdisciplinary center for research and evaluation studies of children and adults in educational settings. The Institute for Child Study is a unit of the School of Education, Indiana University.

The research reports and technical papers are produced under the auspices of the Institute for Child Study and are intended as a channel of communication among the personnel of the Institute for Child Study and students and faculty at Indiana University and elsewhere.

The Institute is also the center of the Reading Clinic, the School psychology program, and the EPDA Project OEG-0-9-247053-3589-721-Proposal To Effect Change in the Training of Teacher Trainers through the Model of Diagnostic Teaching.

Copies of individual papers may be obtained by writing the Director, Institute for Child Study, Indiana University, Bloomington, Indiana 47401, or by communication with individual authors.

Nicholas J. Anastasiow, Ph.D.
Director

Gary Borich, Ed.D.
Evaluator, EPDA Project

Gerald Bracey, Ph.D.
Assistant Director

Susan J. Eklund, Ph.D.
Director, School Psychology

Roger Farr, Ed.D.
Director, Reading Clinic

Alfred Fink, Ph.D.
Special Education, EPDA Project

Kathy Jongsma, M.S.T.
Supervisor, Reading Clinic

Joan Prentice, Ed.D.
Research Associate Institute,
Educational Psychology

Thomas Smith, Ph.D.
Field Coordinator, EPDA Project

Jaap Tuinman, Ph.D.
Research Associate,
Reading Clinic

David Vore, Ph.D.
Research Associate Institute,
School Psychology

Virginia Woodward, Ed.D.
Research Associate Institute,
Early Childhood Education

INSTITUTE: Institute for Child Study, University Schools, Indiana University, Bloomington, Indiana 47401.

ED055740

Increases in Test Scores as a Function of Material Rewards

J. Jaap Tuinman

Roger Farr

B. Elgit Blanton

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Rationale

In the literature surrounding the accountability boom, the issue of raising performance by offering material rewards to students is discussed frequently. Two cases must be distinguished here. First, there is the possibility to entice students to learn more, to study harder, etc. in order to obtain a prize of some kind. Though in a sense achievement gains in this case may be said to be due to manipulation of factors extrinsic to learning, the distinguishing feature of this situation is that high post instructional test scores reflect a genuine increase in the skills or understandings tapped by the test in question.

The second case is characterized by absence of such learning. Here, final test scores are higher than initial best scores merely because of promises of material rewards just prior to taking the test. In this situation achievement scores are raised while it must be assumed that no learning occurred between initial and final measures.

In instructional settings, whenever rewards are available for performance on some post instructional measure, a confounding of both cases exists. It might be hypothesized that performance increases on such a measure may have their source 1) in actual learning, whether or not induced by the awaiting rewards, 2) in an increase of energy and concentration on the post test and 3) in a blend of 1) and 2).

While the above analysis seems plausible enough, little is known

about whether or not post test scores actually can be boosted merely by offering material rewards. The question central to the present study therefore is: Can a significant increase in test scores be induced by offering material rewards in absence of learning?

In view of the increasing emphasis on accountability this question seems to have an undeniable urgency.

A second question asked in this study, related to the first one, is whether such rewards would lead to a higher number of items attempted. This question is of interest since in many cases time limits for standardized reading tests prohibit attempting all items under standard levels of motivation.

Method

Materials

The achievement measure used in this study was form A of the Nelson Reading Test. (Nelson, 1962) This test consists of two parts: vocabulary (100 items) and Comprehension (75 items). The rationale for selecting a test where speed is a relatively large factor is the assumption that increases in energy on the students' part might be most directly reflected in increases in number of the items attempted.

Subjects

In this study a Pre-post randomized group design was utilized. The accessible population was defined as the entire student body in the 7th and 8th grade of a Junior High School in a rural area of Southern Indiana (N=341). From this population the experimental

sample (n=160) was selected randomly, using a random number table. Subjects were then assigned to the Experimental Group and a Control Group by the same random process. (Through an oversight the E-group contained 81 Ss and the C-group 79).

Procedure

At the first testing all Ss in the accessible population were tested. Testing took place in regular class sessions. All 341 students were tested in order to increase the information returns to the school system. By the method mentioned above, the Ss to be retained for post testing (E and C groups) were selected. After an interval of 4 weeks these Ss were post-tested. The C-group was post tested in its entirety in the school cafeteria during the first school period in the morning. Their instructions were simply that they had to take the test again in order to find out how much they had learned sinee the last testing.

Immediately following the testing of the C-group, the E-group was tested. The treatment for this group consisted of the availability of performance rewards: 6 transitor radios (at $4.95), 9 sweatshirts with the emblem of nearby Indiana University and a supply of 25 cent candy bars. The directions for this group were:

> "A few weeks ago you took a test and we have now scored your test. We would like to have you try to beat your first test score. We have some prizes for you if you can do better on this than you did on the last test. We don't care who makes the highest score. We want to see who can make <u>more points</u> on this test than on the first test. All you need to do is <u>raise your own score</u> by as many points as you can. The students

who raise their scores the most will win prizes. Remember, it's not the student that scores highest that wins; it's the student tho raises his score the most.

If you are one of the top 6 students who gain the most points you will win a transitor radio --- NOT the 6 highest scores but the 6 who improve the most.

If you are one of the next 9 students who raise their score the most points, you will win one of these IU shirts.

And if you raise your score even as little as one point, you get one of these candy bars. So everyone who gets even just one point higher will get a prize.

If you do not score any higher on this test than the first test there are no prizes.

It's not hard, just try to beat your own score. Compete with no one but yourself. All you have to do is work harder and try to do more items and more of them correctly than you did last time. Raise your score and take home one of the prizes."

Analysis

The results were analyzed using analysis of covariance, with pre-test scores as the covariate in the case of all dependent variables. Two major dependent variables were subjected to analysis: the total numbers correct, (NC) and the total number of items attempted (NA). Results for the vocabulary and comprehension subtests are also reported.

Results

Table 1 summarizes the descriptive data for the number correct (NC) dependent variable. Since the slopes of the within group repression lines were not significantly different, in fact, they were nearly identical, ANCOVA was used as planned. Table 2 shows

post test means adjusted for pretest differences between the experimental and control group mean scores. As was to be expected the coefficients of homogeneity were quite high (Table 3) matching those reported by the test publisher. It may be mentioned that the stability coefficients (product moment correlation between pre and post-test scores) were somewhat lower. For vocabulary, comprehension and total scores these coefficients were for the Experimental group .82, .74 and .85 respectively; for the Control Ss these correlations were .73, .84 and .83.

In Tables 4 through 6 the results of the ANCOVA on the NC variable are reported. The effect of the treatment is significant in all instances. The probability that differences on the total score variable could have arisen by chance rather than as a consequence of experimental manipulation is smaller than .01.

As was discussed above, it was assumed that the Experimental Ss would attempt more items than the Control group. It must be pointed out that a test on the difference between mean number of items attempted is not a direct test on the hypothesis that the increased number correct is a function of the increase in numbers attempted. The notion that increase in speed led to more errors but that these errors were offset by an increase in concentration on the experimental subject's part was not tested in this experiment. The correlation coefficients in Table 7 seem to suggest a careful interpretation of the relation between number of attempts and number of items correct.

Table 8 summarizes the descriptive data for the dependent variable Number of items Attempted (NA). In Table 9 post-test means adjusted for pre-test differences are presented.

From Tables 10-12 it can be seen that the differences between treatment means when NA is the dependent variable, is significant at the .001 level. The experimental group attempted significantly more items that the control group.

Table 13 summarizes the ANCOVA's and indicates the level of significance of the various effects.

Discussion

The data presented above indicates unequivocally that it is possible to raise test scores merely by presenting material rewards.

Again it must be pointed out that the differences between experimental and control group means cannot be explained by reference to any learning.

Here is a case where differences not only have statistical significance but practical significance as well. According to the table of norms in the publisher's manual, the experimental group out performed the control group by more than 3 months (based on adjusted total means, Table 2).

From the point of view of performance contracting business, the results obtained are quite interesting. Figure 1 shows a table of payments which is based on a real contract ( ). In this contract gains per year teaching are expressed in portions

of a year. For the purpose of this study, the authors used the same base for payments but substituted the unit "year" for "month", since the experiment lasted only 4 weeks. On this basis the author "earned" $5,260, if no penalty was enforced for students who dropped back, or $2,860 with such a penalty. No small return for an investment of six radios, nine sweaters and some candy, or $75 in all.

Table 1

Means and Standard Deviations for the Experimental (n=81) and the Control Groups. (n=79) (Number correct is dependent variable)

	Pre-test			Post-test		
	Vocab.	Compre.	Total	Vocab.	Compre.	Total
Experimental (n=81)	56.20 (10.56)	42.00 (11.02)	98.20 (20.73)	62.37 (11.88)	50.62 (12.04)	112.99 (22.91)
Control (n=79)	54.91 (11.06)	(40.46) (11.42)	95.37 (21.08)	(58.71) (12.69)	46.87 (13.07)	105.60 (24.91)

Table 2

Means and Standard Errors on the Post-test adjusted for Pre-test differences (Number Correct is Dependent Variable)

Group	Vocabulary	Comprehension	Total
Experimental	61.71 (.54)	49.88 (.68)	111.51 (.96)
Control	59.39 (.55)	47.65 (.69)	107.13 (.98)

Table 3

Reliability Coefficients and Standard Errors of Measurement

			KR20	$S.E._m$	Spearman Brown (Split half)	$S.E._m$
Pretest		Vocabulary	.94	2.62	.94	2.27
		Comprehension	.93	2.73	.97	1.59
Post Test	Experimental	Vocabulary	.95	2.75	.97	2.03
		Comprehension	.95	2.99	.96	2.49
	Control	Vocabulary	.95	2.74	.97	2.23
		Comprehension	,95	2.80	.97	2.10

Table 4

ANCOVA for Vocabulary Post test Scores. Pretest Vocabulary scores as covariate

(Number Correct is Dependent Variable)

Source	SS	df	MS	F
Treatment	215.37	1	215.37	8.98**
Within	3791.57	158	24.00	
Total	4006.94			

**p < .01

Table 5

ANCOVA for Comprehension Post Test Scores

Pretest Vocabulary Scores as Covariate (Number Correct is Dependent Variable)

Source				
Treatment	198.74	1	198.74	5.28*
Within	5947.63	158	37.64	
Total	6146.37			

*p < .05

Table 6

ANCOVA for Total Reading Post Test Scores

Pretest Total Reading Scores as Covariate

(Number Correct is Dependent Variable)

Source				
Treatment	767.55	1	767.55	10.16**
Within	11931.25	158	75.51	
Total	12698.80			

**p < .01

Table 7

Pearson Product Moment Correlations between

Number Correct and Number Attempted Scores

	Pre-test			Post-test		
	Voc.	Compreh.	Total	Voc.	Compreh.	Total
Experimental (n=81)	.88	.83	.88	.88	.75	.84
Control (n=79)	.76	.78	.78	.90	.84	.89

Table 8

Means and Standard Deviations for the Experimental and Control Groups

Number Attempted is the Dependent Variable

	Pretest			Post Test		
	Voc.	Compreh.	Total	Voc.	Compreh.	Total
Experimental (n=82)	62.02 (10.13)	48.01 (10.60)	110.03 (19.20)	72.21 (13.32)	60.12 (11.63)	132.33 (23.29)
Control (n=79)	62.49 (12.07)	47.72 (10.63)	110.22 (20.99)	66.52 (12.06)	54.23 (12.31)	120.75 (22.95)

Table 9

Means and Standard Errors on the Post Test, adjusted for Pretest differences. (Number Attempted is Dependent Variable)

	Vocabulary	Comprehension	Total
Experimental (n=81)	72.41 (.91)	59.99 (.82)	132.40 (1.42)
Control (n=79)	66.31 (.92)	54.36 (.84)	120.66 (1.44)

Table 10

ANCOVA for Vocabulary Post Test Scores.

Pretest Vocabulary Score as Covariate. (Number Attempted is Dependent Variable)

Source	SS	df	MS	F
Treatment	1495.75	1	1495.75	22.28***
Within	10606.56	158	67.13	
Total	12102.31			

*** p< .001

Table 11

ANCOVA for Comprehension Post Test Scores.

Pretest Comprehension scores as covariate.

(Number Attempted is Dependent Variable)

Source	SS	df	MS	F
Treatment	1275.04	1	1275.04	22.89***
Within	8802.04	158	55.71	
Total	10077.08			

*** $p < .001$

Table 12

ANCOVA for Total Reading Post Test Score.

Pretest scores as covariate. (Number attempted is dependent variable.)

Source	SS	df	MS	F
Treatment	5560.18	1	5560.18	33.82***
Within	25975.43	158	164.40	
Total	31535.61			

*** $p < .001$

Table 13

Analyses of Covariance for Items Correct and Items Attempted Post Test mean scores, adjusted for their respeciive Pretest mean scores. Table entries are F-values.

	Vocabulary	Comprehension	Total
Correct	8.98**	5.28*	10.16**
Attempted	22.28***	22.89***	33.82***

* $p < .05$

** $p < .01$

*** $p < .001$

Figure 1

Table of Payments

Schedule of Payments:

Gain	Pay
0.10 -- .19 mths	$100.00
0.20 -- .29 mths	$130.00
0.30 -- .39 mths	$140.00
0.40 -- + mths	$150.00

To be paid to the experimental students after subtraction of mean gain of control group from each individual's gain.

A	B
(without penalty for loss)	(with penalty for loss)
$5260	$2860

REFERENCES

1. M. J. Nelson, The Nelson Reading Test. Boston: Houghton Mifflin Comp., 1962.

2. Performance Contract between the School District of Grand Rapids, Michigan and Alpha Learning System Company. East Lansing, Mich.: Michigan Education Association, December, 1970.